Abstract

We have compared the performance of four approaches for automatic language identification of speech utterances: Gaussian Mixture Model (GMM) classification; single-language Phone Recognition followed by language-dependent, interpolated n-gram Language Modeling (PRLM); parallel PRLM, which uses multiple single-language phone recognizers, each trained in a different language; and language-dependent Parallel Phone Recognition (PPR). These approaches, which span a wide range of training requirements and levels of recognition complexity, were evaluated with the Oregon Graduate Institute Multi-Language Telephone Speech Corpus. Systems containing phone recognizers performed better than the simpler GMM classifier. The top-performing system was parallel PRLM, which exhibited an error rate of 2% for 45-s utterances and 5% for 10-s utterances in two-language, closed-set, forced-choice classification. The error rate for 11-language, closed-set, forced-choice classification was 11% for 45-s utterances and 21% for 10-s utterances.